



Markup UK 2023

Improving quality-critical XML workflows
with XProc 3.0 pipelines

Achim Berndzen, `<xml-project />`

Thorsten Rohm, Thieme Compliance GmbH

Improving quality-critical XML workflows with XProc 3.0 pipelines

- Orchestrating complex XML pipelines has been a major topic of XML related software development over the years.
- Comprehensive techniques have been developed to
 1. Deliver high-quality results
 2. Ensure that the pipelines can be maintained
 3. Allow the pipelines to be debugged for straightforward troubleshooting
- Quality demands for workflows and results might vary.
- Implementing new quality demands for results might have negative impacts on pipelines quality.
- Then it might be time to start all over again...

Improving quality-critical XML workflows with XProc 3.0 pipelines

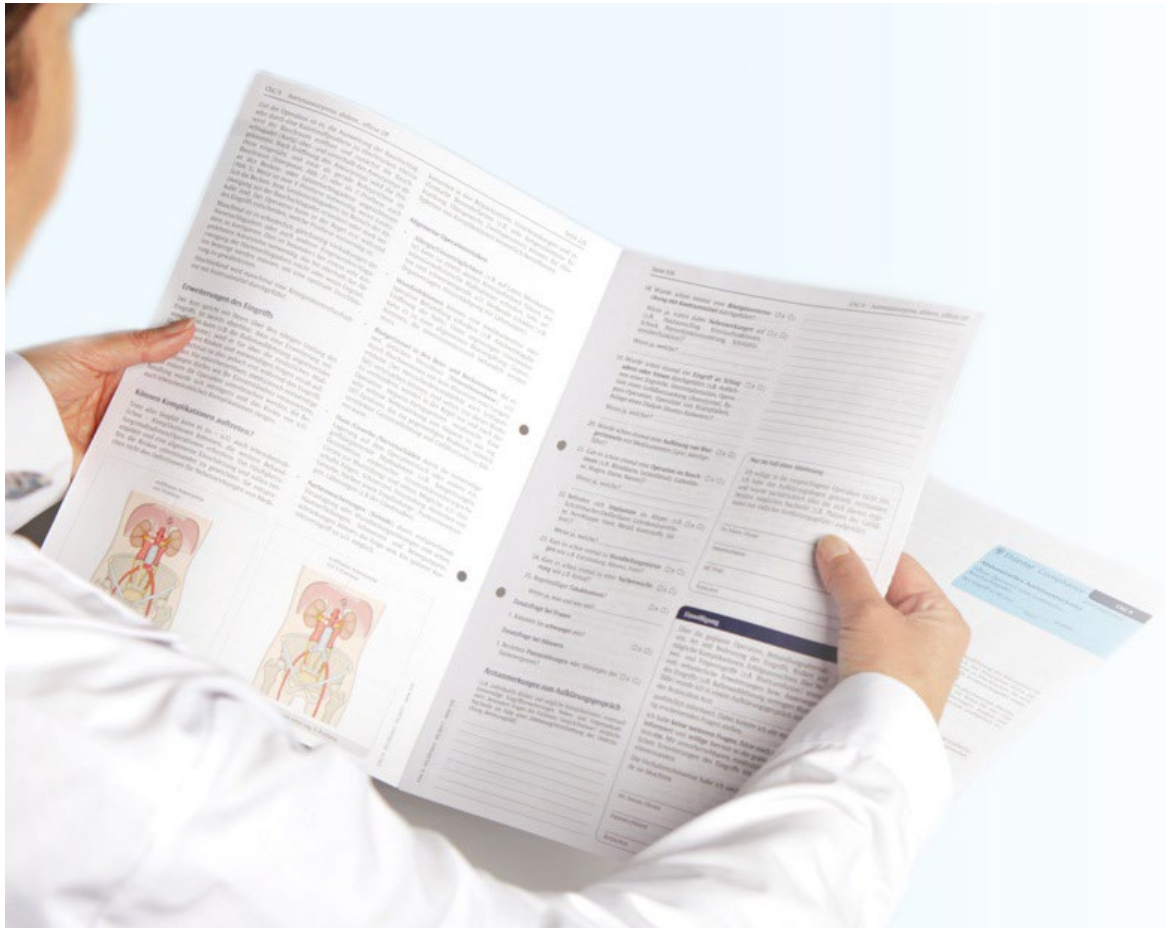
- Shared project of our two companies.
- Adding new quality features to well established pipelines producing documents with high-quality demands.
- The existing workflow already had some pain points.
- Taken together: We decided to start over and refactor the workflow.
- And we even decided to change the basic orchestration technology:

Windows batch files
with XSLT stylesheets



XProc 3.0 based on the XSLT stylesheets,
doing away shell scripting as much as possible.

Overview



© Thieme

- Background
 - About Thieme Compliance GmbH and patient education leaflets
 - A very short view on XProc 3.0
- Introducing the existing batches
- Pain points of the existing batches
- New requirements for next version
- New system based on XProc 3.0
- Takeaways



Can you judge a book by its cover?

In the case of Thieme, we think you can ...



Background: About Thieme Compliance GmbH and patient education leaflets

From German-speaking market-leader to global medical publishing specialists

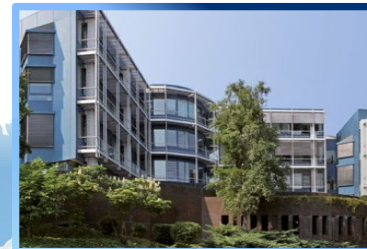
Neurosurgery, radiology, plastic surgery, and beyond – people worldwide come to Thieme for expert insights



New York since 1979



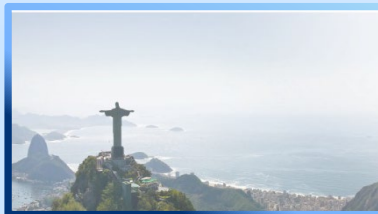
Germany since 1886



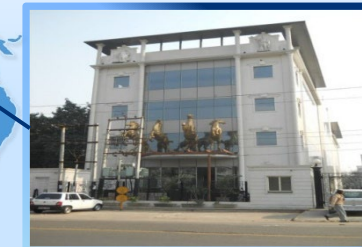
Beijing since 2018



Rio de Janeiro since 2013



Delhi since 2008

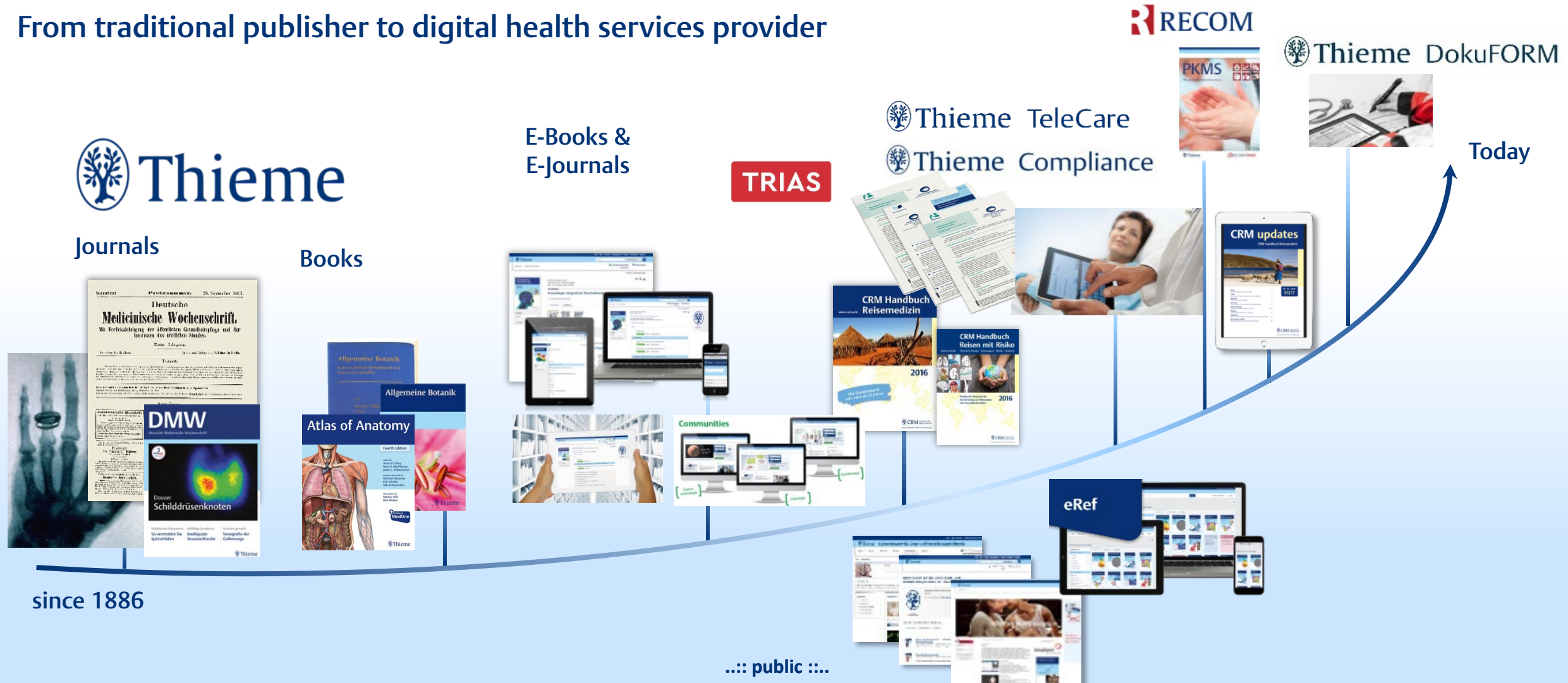


11 locations worldwide,
1,000 international employees
and numerous sales partners
all over the world

...: public ...

Background: About Thieme Compliance GmbH and patient education leaflets

From traditional publisher to digital health services provider



Background: About Thieme Compliance GmbH and patient education leaflets



THIEME'S MISSION:

To improve health and healthcare – 

by providing key information
at the right time and the right place.

For better medicine and a healthier life.



Background: About Thieme Compliance GmbH and patient education leaflets

Patient education leaflets

- Patient education is an important aspect of healthcare and an important legal and ethical principle in medicine.
- Thieme Compliance GmbH supports healthcare facilities in implementing this principle by providing customised information materials tailored to the specific needs of patients.
- The aim is to help ensure that patients are better informed and educated so that they can make decisions about their health in close cooperation with their doctors.

Thieme Compliance A1-6 GB
Anaesthesia in Adults and Adolescents
(General Anaesthesia and/or Regional Anaesthesia)

Dear Patient, Dear Parents,
This informed consent form is intended to prepare you for the patient-doctor discussion. Please read it carefully and complete the questionnaire carefully and completely. For better readability, we use male pronouns but are addressing all genders with them.

Which anaesthetic procedures are possible?
To ensure that you will feel no pain during the proposed procedure (examination, treatment or operation), it will be performed under anaesthesia. Pain can be eliminated by means of **general anaesthesia** and/or **regional anaesthesia**. The doctor will discuss with you which type of anaesthesia is best suited for you, which other anaesthetic procedures are also an option and the advantages, disadvantages and risks of the individual methods.

Before the anaesthesia, an indwelling catheter is placed in the vein of your hand or arm. It is used to administer sedatives, pain medications, other medications (e.g. antibiotics) and infusions as needed.

General anaesthesia
Under general anaesthesia, you are in deep sleep in which consciousness and pain perception are temporarily eliminated. For the induction of general anaesthesia, the doctor injects a fast-acting anaesthetic agent via the indwelling catheter. In order to maintain the anaesthesia, you receive

- further anaesthetic agents via the indwelling catheter (intravenous anaesthesia) or
- anaesthetic gas that is administered together with respiratory air.

Very often, both procedures are combined. Your doctor ensures your supply of oxygen and, as required, anaesthetic gases by means of

- a **breathing mask**, which is placed over the mouth and nose, or
- a **mask placed over the larynx (laryngeal mask)**, which is inserted through the mouth and placed over the laryngeal inlet, or
- a **breathing tube**, which is inserted into the trachea.

The doctor will insert the **laryngeal mask** or **breathing tube** when you are already asleep. Both of these enable artificial ventilation and protect the airways. The laryngeal mask and the breathing tube in particular reduce the risk of aspiration of saliva and stomach contents into the lungs. In order to gently insert the breathing tube, the doctor administers a **muscle relaxant**. For the insertion of the laryngeal mask, this is necessary in exceptional cases only.

Regional anaesthesia
Regional anaesthesia eliminates any pain in the operated part of the body. The doctor injects a local **sedative** close to the pain-conducting nerves going towards and supplying the operation site. The agent numbs the nerves so that pain stimuli cannot be transmitted to the brain (**nerve block**). Usually, this eliminates the pain sensation for several hours. After the injection of the sedative, you will feel some **tingling and warmth**. The sedated body regions (e.g. arm, hand, lower body, legs) then become **numb**, and as long as the effect continues, they **cannot be moved** or can be moved to a **limited extent only**.

If **regional anaesthesia** is used solely, you will be awake and responsive. However, you may also receive a **sedative** or **sleep-inducing medication (sedation)**. In this case, you will have only a **limited recollection** of the procedure or no recollection at all.

Regional anaesthesia is frequently used in addition to general anaesthesia. The advantage for the patients is that they need less anaesthetics, recover faster from the procedure and experience little pain directly after the operation. However, the additional regional anaesthesia is also associated with additional risks.

For procedures on the arm or hand, the nerve plexus of the arm is sedated. This is achieved by injecting a sedative into the arm pit (Fig. 1). It is the site of nerves running along the arm up to the fingers. Before the injection, the correct position of the tip of the needle close to the nerves is determined by the doctor with the aid of ultrasound or a nerve stimulator. In the short term, nerve stimulation can cause muscle twitching and unpleasant sensations ("electrification").

Epidural anaesthesia and spinal anaesthesia
These types of anaesthesia are used for procedures on the breast/chest, abdomen, back, pelvis, groin and legs. The doctor injects an anaesthetic agent into the region of the spine to sedate pain-conducting nerves emerging from the spinal cord.

- **Epidural anaesthesia:**
The doctor inserts a hollow needle into the **epidural space** in the back and threads a thin synthetic tube through it (Fig. 2). The needle is then removed, and the anaesthetic agent is injected through the catheter. Depending on the type of surgical procedure needed, the anaesthetic agent can be injected at the level of the thoracic spine (thoracic epidural anaesthesia) or the lumbar spine (lumbar epidural anaesthesia).
- **Spinal anaesthesia:**
Through the dura mater, the doctor inserts a thin needle into the **spinal canal**, which is filled with cerebrospinal fluid, in the region of the lumbar spine (Fig. 2). He then injects the anaesthetic agent into this canal.

In isolated cases, **spinal anaesthesia and epidural anaesthesia** can also be combined with each other.

Further measures
Usually, pain is eliminated very well by regional anaesthesia. However, the effect of the regional anaesthesia can be **insufficient** in isolated cases. The doctor can then inject **further sedatives and pain medications**. If this is impossible, if the anaesthesia spreads too far or if complications occur, the procedure must be continued under **general anaesthesia**. The doctor can place a **catheter** (epidural, spinal or brachial plexus catheter) for pain management after the procedure or can leave a catheter in place which is used for anaesthesia. If this applies in your case, you will be counselled about it in a separate patient-doctor discussion. Your doctor will also discuss with you any foreseeable **additional measures** (e.g. placement of a central venous catheter) in a separate patient-doctor discussion. If a **transfusion of blood from a foreign donor** is also seriously considered, you will be informed about how it is performed, its risks (e.g. hypersensitivity/incompatibility reactions, infection, e.g. hepatitis or HIV infection [with AIDS as a late consequence] in extremely rare cases, under certain circumstances also infection with unknown pathogens) as well as any measures to avoid using foreign donor blood in a separate patient-doctor discussion.

Risks and possible associated complications
The doctor ensures your safety by monitoring your vital body functions (e.g. pulse, blood pressure, breathing) during the anaesthesia and supporting them if necessary (e.g. by administering circulation medications). However, despite the greatest care taken, complications can arise which can even become life-threatening and necessitate additional treatment or further surgery under certain circumstances. The frequency rates are only a general estimate and are intended for weighing the risks against each other. They are not the same as the definitions of side-effects stated in the package inserts of medications. Pre-existing underlying diseases and individual unusual circumstances can significantly influence the rate of complications.

In the following, you are provided with information on the risks with which the anaesthetic procedures are generally associated, even if some complications only occur very rarely or temporarily.

If the administration of medications which have been proven to be successful in anaesthesia but do not have formal approval (**off-label use**) is planned for you, the doctor will discuss this with you and inform you about the known risks. However, unknown risks cannot be ruled out. Under certain circumstances, the manufacturer may also not accept any liability.

General risks

- **Injury to blood vessels** can be caused by the tips of the needles, cannulas or catheters. This can cause **bleeding** and **haematomas**. Treatment or an operation is necessary in rare cases only.

Fig. 1: Injection area for the brachial plexus anaesthesia

Fig. 2: Injection area for the spinal/epidural anaesthesia

Background: About Thieme Compliance GmbH and patient education leaflets

Patient education leaflets

- The content is developed in close cooperation with more than 400 experts from the medical community and tested for their comprehensibility and usefulness.
- A team of legal advisors ensures that the content always corresponds to current case law.
- In total, more than 2,000 patient education leaflets from more than 30 speciality areas are available in up to 31 languages.
- The leaflets are available in various formats in digital form as well as print.



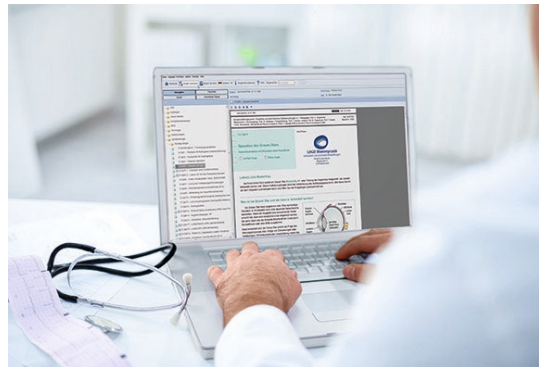
Background: About Thieme Compliance GmbH and patient education leaflets

Patient education leaflets enhanced with services facilitate patient communication and education and support processes in clinics and practices

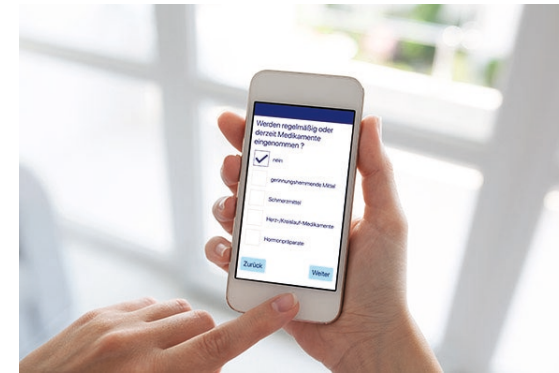
Patient education



Process documentation



Digital workflow



**E-ConsentPro
mobile**

- Saves time through location-independent patient information
- Optimizes processes with digital and paperless workflows
- Excellent patient satisfaction thanks to high-quality multimedia information
- Better compliance on the part of the patient
- Improves use of time and resources

...: public ...

Background: About Thieme Compliance GmbH and patient education leaflets

- Syntactic and semantic healthcare interoperability of patient education leaflets
- For the medical history part of the patient education leaflets, HL7's FHIR (Fast Healthcare Interoperability Resources) questionnaire resource is used for syntactic healthcare interoperability.
- To ensure semantic healthcare interoperability too, the questionnaire resource contains codings from SNOMED CT (Systematized Nomenclature of Medicine and Clinical Terms) or LOINC (Logical Observation Identifiers Names and Codes).

```
<fhir:item>
  <fhir:linkId value="MF_Erkrankungen_Familie_Erkrankung_Blutsverwandtschaft" />
  <fhir:text value="Among your blood relatives, are there any diseases or indications of a disease?" />
  <fhir:type value="open-choice" />
  <fhir:required value="true" />
  <fhir:repeats value="true" />
  <fhir:answerOption id="MF_Erkrankungen_Familie_Erkrankung_Blutsverwandtschaft_nein">
    <fhir:extension url="http://hl7.org/fhir/StructureDefinition/questionnaire-optionExclusive">
      <fhir:valueBoolean value="true" />
    </fhir:extension>
    <fhir:valueCoding>
      <fhir:system value="http://snomed.info/sct" />
      <fhir:version value="http://snomed.info/sct/900000000000207008/version/20220430" />
      <fhir:code value="160266009" />
      <fhir:display value="No family history of clinical finding (situation)" />
    </fhir:valueCoding>
  </fhir:answerOption>
  <fhir:answerOption id="MF_Erkrankungen_Familie_Erkrankung_Blutsverwandtschaft_Krebs">
    <fhir:valueCoding>
      <fhir:system value="http://snomed.info/sct" />
      <fhir:version value="http://snomed.info/sct/900000000000207008/version/20210131" />
      <fhir:code value="275937001" />
      <fhir:display value="Family history of cancer (situation)" />
    </fhir:valueCoding>
  </fhir:answerOption>
  <!-- ... -->
  <fhir:answerOption id="MF_Erkrankungen_Familie_Erkrankung_Blutsverwandtschaft_Erbkrankheiten">
    <fhir:valueCoding>
      <fhir:system value="http://snomed.info/sct" />
      <fhir:version value="http://snomed.info/sct/900000000000207008/version/20220430" />
      <fhir:code value="429962007" />
      <fhir:display value="Family history of hereditary disease (situation)" />
    </fhir:valueCoding>
  </fhir:answerOption>
</fhir:item>
```

...: public ...

Background: A very short view on XProc 3.0

- XProc 3.0: A pipeline language with an XML based syntax:
 - Derived from XProc (W3C recommendation from 2010).
 - Developed by a XML community group of volunteers.
 - Published in two community reports in September 2022.
- Highlights:
 - Basic document model with JSON, HTML, text and binary document in addition to XML.
 - XPath 3.1 as basic expression language, XDM typing for variables and options.
 - Enlarged standard step library: Steps for archives, text documents, etc.
 - Lot of syntactic sugar to improve smoother coding experience.

Background: A very short view on XProc 3.0

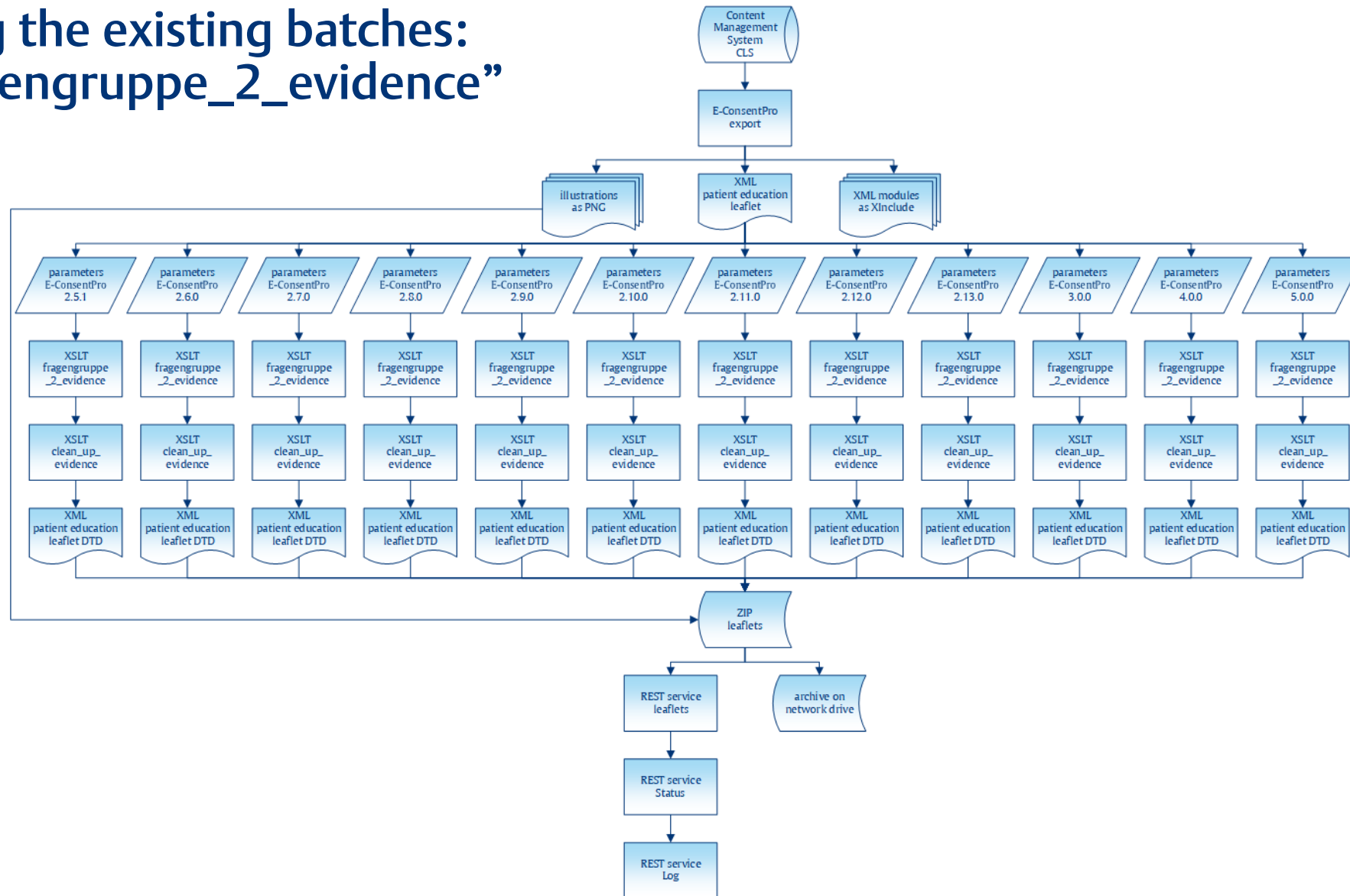
Better documentation:

- <https://www.xproc.org> provides an overview over learning material.
- Highlight: Erik Siegel's The XProc 3.0 Programmer Reference.

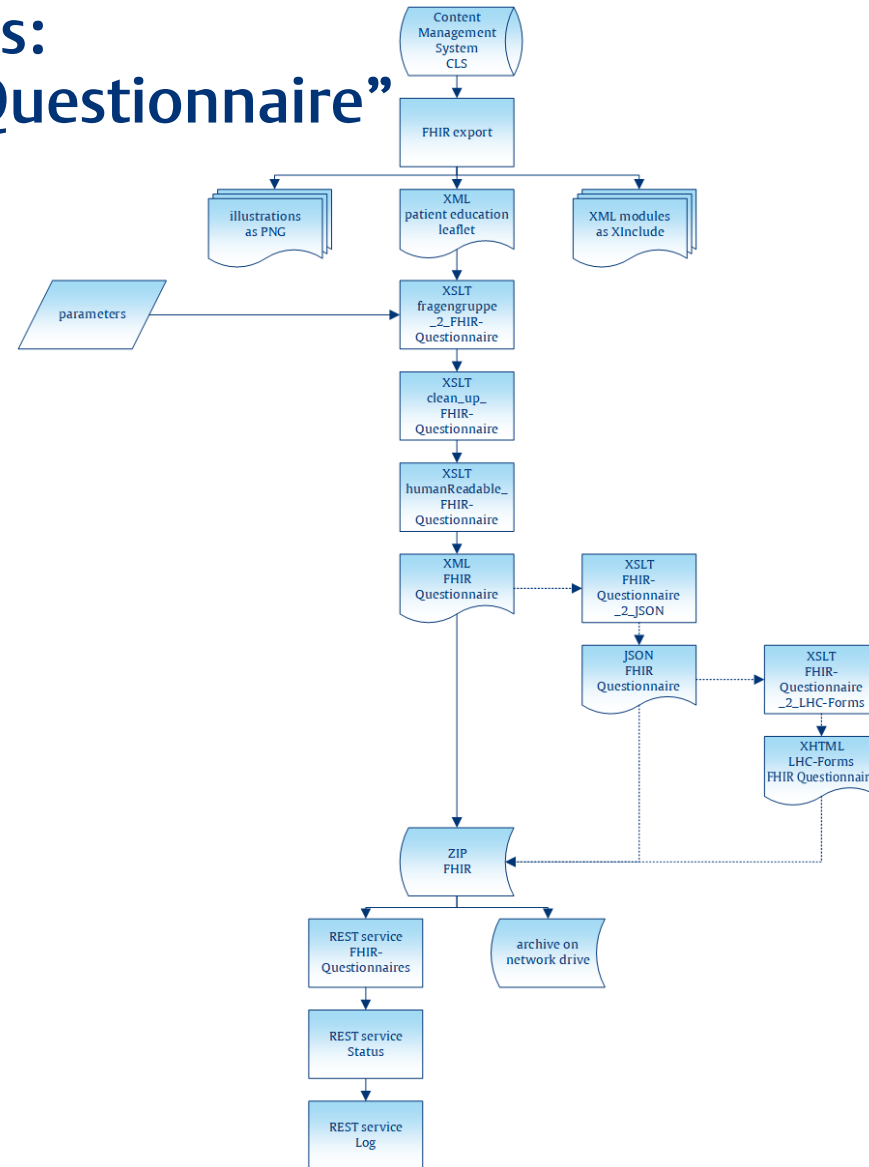
Two Implementations:

- XML Calabash 3 by Norm Tovey-Walsh, successor to XML Calabash, de facto standard for XProc 1.0
- MorganaXProc-IIIse by <xml-project /> is used in this project
 - Implements all required and most optional features of the core specification.
 - Additional to the standard step library, file- and validation steps are supported.
 - Public beta since February 2020, Version 1.0 was released in September 2022.
 - Passes all relevant tests in the XProc 3.0 test suite.
 - Open source product released under GNU GPL 3.0.

Introducing the existing batches: Batch „fragengruppe_2_evidence“



Introducing the existing batches: Batch „fragengruppe_2_FHIR-Questionnaire”



...: public ...

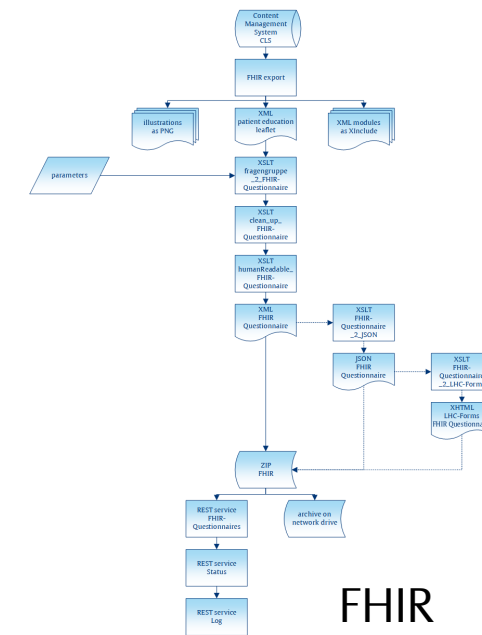
Pain points of the existing batches: Summary

- Lacking of flexibility for inserting additional XSLT steps (in between)
- No easy way to debug the intermediate results of each XSLT step
- Too many tools means too many dependencies: 7-Zip, curl, Beyond Compare

New requirements for version next: Summary

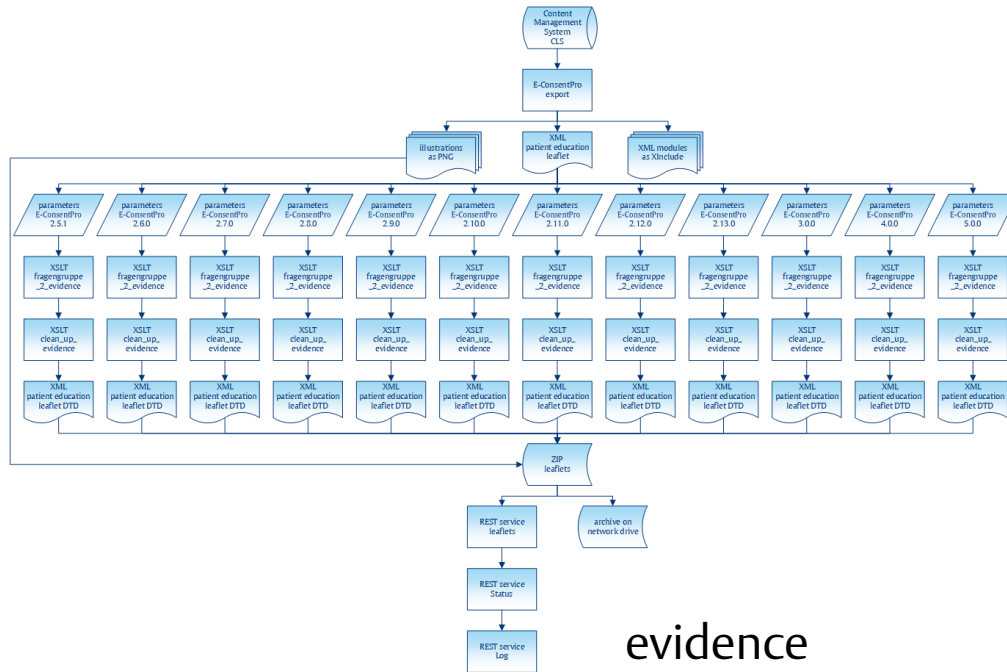
- Increased quality through validation of XML sources using T_0 XSD as well as validation of XML results using specific versions of T_0 DTD
- Increased quality by additional validation of XML results using Schematron
- Summarised, formatted and easily comprehensible log files
- Performance improvement by omitting unnecessary images from the Zip archive
- Limiting processing to specific sources from the source folder

New system based on XProc 3.0



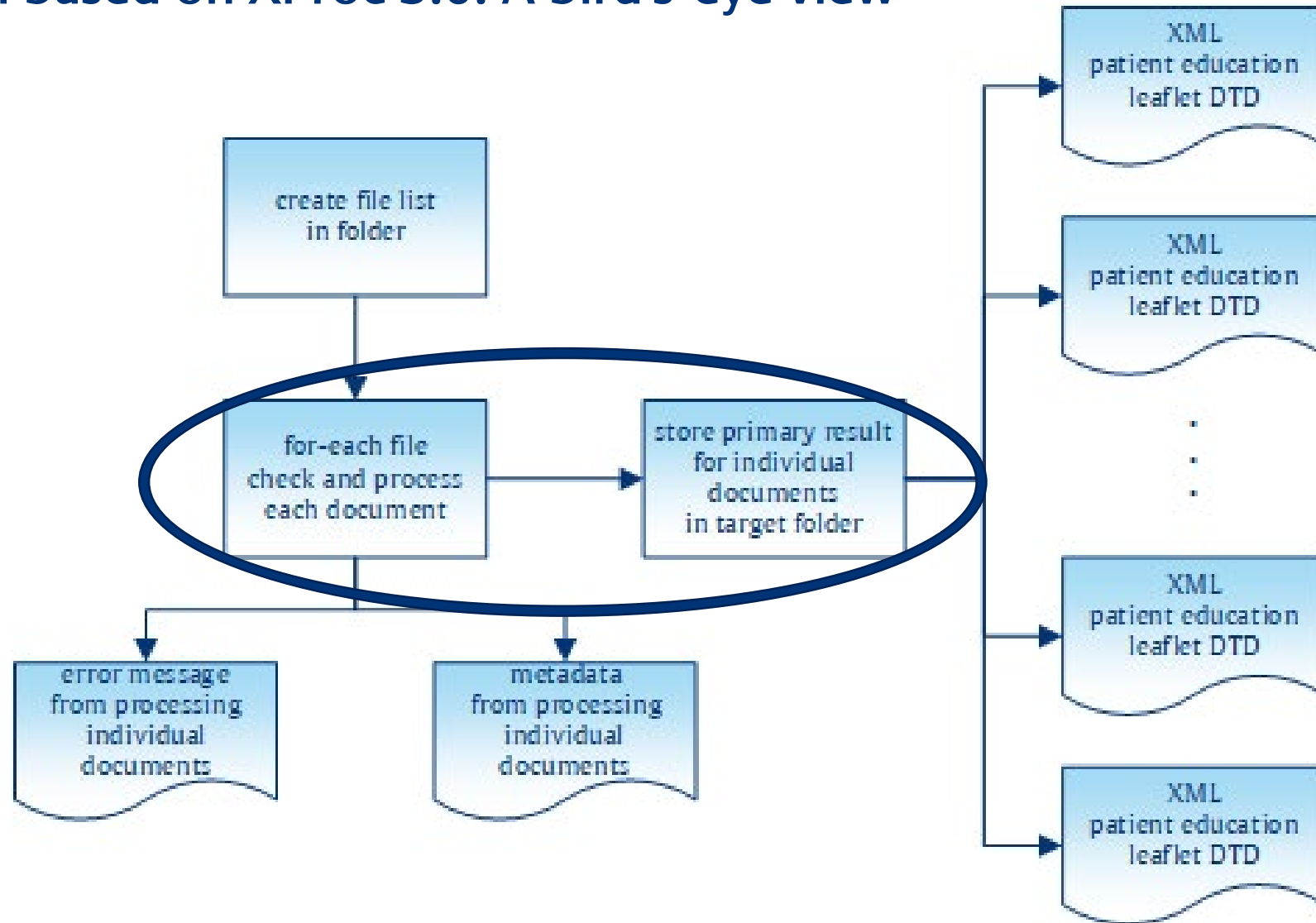
- Refactor both workflows as XProc 3.0 pipelines
- Remove batch processing, keep the elaborated and well-tested XSLT stylesheets
- Replace third party software with XProc steps

New system based on XProc 3.0



- Original workflow has double flow-control:
 - Twelve sub-batches called for every result version.
 - Each batch calls XSLT iterating over all source documents.
- Replace it with one XProc 3.0 control flow.
- Main reason comes from new requirements:
 - Cross-reference check for images
 - Perform XSLT transformations only if test is passed.
- Setback: Loosing Saxon's CLI features for XInclude and validation. Replace them with XProc steps.

New system based on XProc 3.0: A bird's-eye view



...: public ...

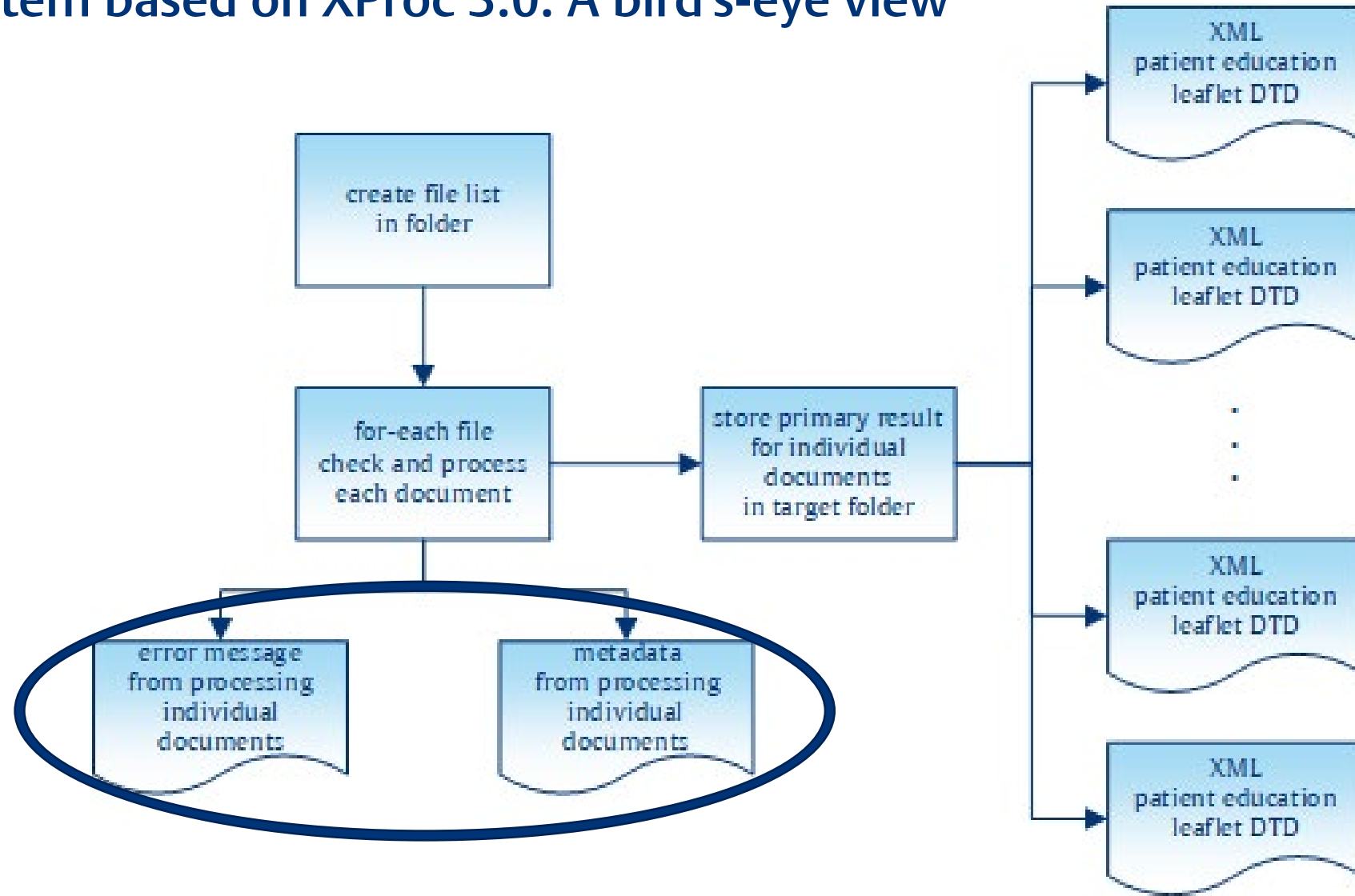
The new system based on XProc 3.0: Iterating over all documents

- Up to six XSLT stylesheets were connected by Saxon's `@saxon:next-in-chain` on `<xsl:output>`.
- Replaced by consecutive calls of `<p:xslt>` where the later step is automatically connected to output of previous step.
- Inserted `<p:store>` between two `<p:xslt>` to write transformation output to disk before processing with next XSLT.
- Non-disruptive change in XProc 3.0: Input document appears on output port. Useful change compared to XProc 1.0!
- Even more useful with `@use-when`: New static options allow to switch debugging on and off from pipeline invocation.

The new system based on XProc 3.0: Iterating over all documents

- Validating final results: Just added `<p:validate-with-xml-schema>` and `<p:validate-with-schematron>` after last XSLT transformation.
- Validation with DTD is a bit more tricky.
- Extremely useful: New `<p:catch code="XXXX" />` makes it easy to create detailed error reports.

New system based on XProc 3.0: A bird's-eye view



The new system based on XProc 3.0: Processing iteration results

```
<tcg:report file="reference-to-source-doc">
  <tcg:report-done phase="validation">
    <successfully-validated />
  </tcg:report-done>
  <tcg:report-done phase="processFileRefs">
    <c:entry found-in="reference-to-source-doc" href="path-to-pic1"
      name="name-of-zip-entry-for-pic1" />
  </tcg:report-done>
  <tcg:report-done phase="2.5.1">
    <c:entry derived-from="reference-to-source-doc" href="path-to-doc1"
      name="name-of-zip-entry-for-doc1" />
  </tcg:report-done>
  <tcg:report-error phase="2.6.0">
    <c:errors><!-- detailed error report here --></c:errors>
  </tcg:report-error>
</tcg:report>
```

The new system based on XProc 3.0: Processing iteration results

Creating the Zip archive:

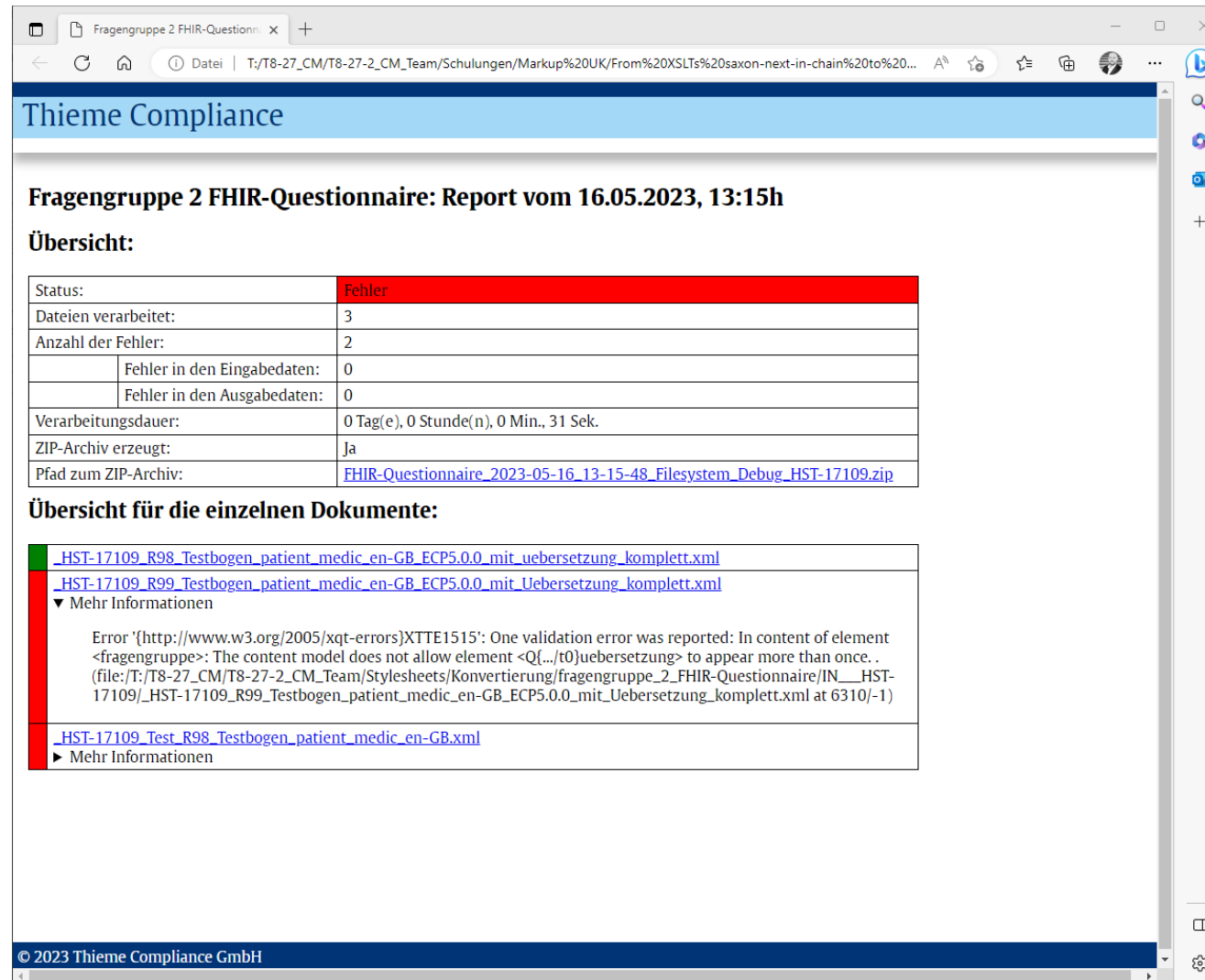
- A `<c:entry>` element is created
 - for every referenced image in a document
 - for every valid result of a transformation sequence
- Some post processing required: Images may come from source not delivering a valid results.
- Creating the Zip archive at the end is enabled by XProc's standard step `<p:archive>`.

The new system based on XProc 3.0: Processing iteration results

Creating the report:

- For every source document there is a `<tcg:report>` element reporting about the different phases.
- Aggregate number of processed documents and found errors.
- Create final HTML report document using another XSLT.

The new system based on XProc 3.0: Processing iteration results



Thieme Compliance

Fragengruppe 2 FHIR-Questionnaire: Report vom 16.05.2023, 13:15h

Übersicht:

Status:	Fehler
Dateien verarbeitet:	3
Anzahl der Fehler:	2
Fehler in den Eingabedaten:	0
Fehler in den Ausgabedaten:	0
Verarbeitungsdauer:	0 Tag(e), 0 Stunde(n), 0 Min., 31 Sek.
ZIP-Archiv erzeugt:	Ja
Pfad zum ZIP-Archiv:	FHIR-Questionnaire_2023-05-16_13-15-48_Filesystem_Debug_HST-17109.zip

Übersicht für die einzelnen Dokumente:

- [_HST-17109_R98_Testbogen_patient_medec_en-GB_ECP5.0.0_mit_uebersetzung_komplett.xml](#)
- [_HST-17109_R99_Testbogen_patient_medec_en-GB_ECP5.0.0_mit_Uebersetzung_komplett.xml](#)
 - ▼ Mehr Informationen
 - Error '[http://www.w3.org/2005/xqt-errors]XTTE1515': One validation error was reported: In content of element <fragengruppe>: The content model does not allow element <Q{.../t0}uebersetzung> to appear more than once. . (file:/T:/T8-27_CM/T8-27-2_CM_Team/Stylesheets/Konvertierung/fragengruppe_2_FHIR-Questionnaire/IN___HST-17109/_HST-17109_R99_Testbogen_patient_medec_en-GB_ECP5.0.0_mit_Uebersetzung_komplett.xml at 6310/-1)
- [_HST-17109_Test_R98_Testbogen_patient_medec_en-GB.xml](#)
 - Mehr Informationen

© 2023 Thieme Compliance GmbH

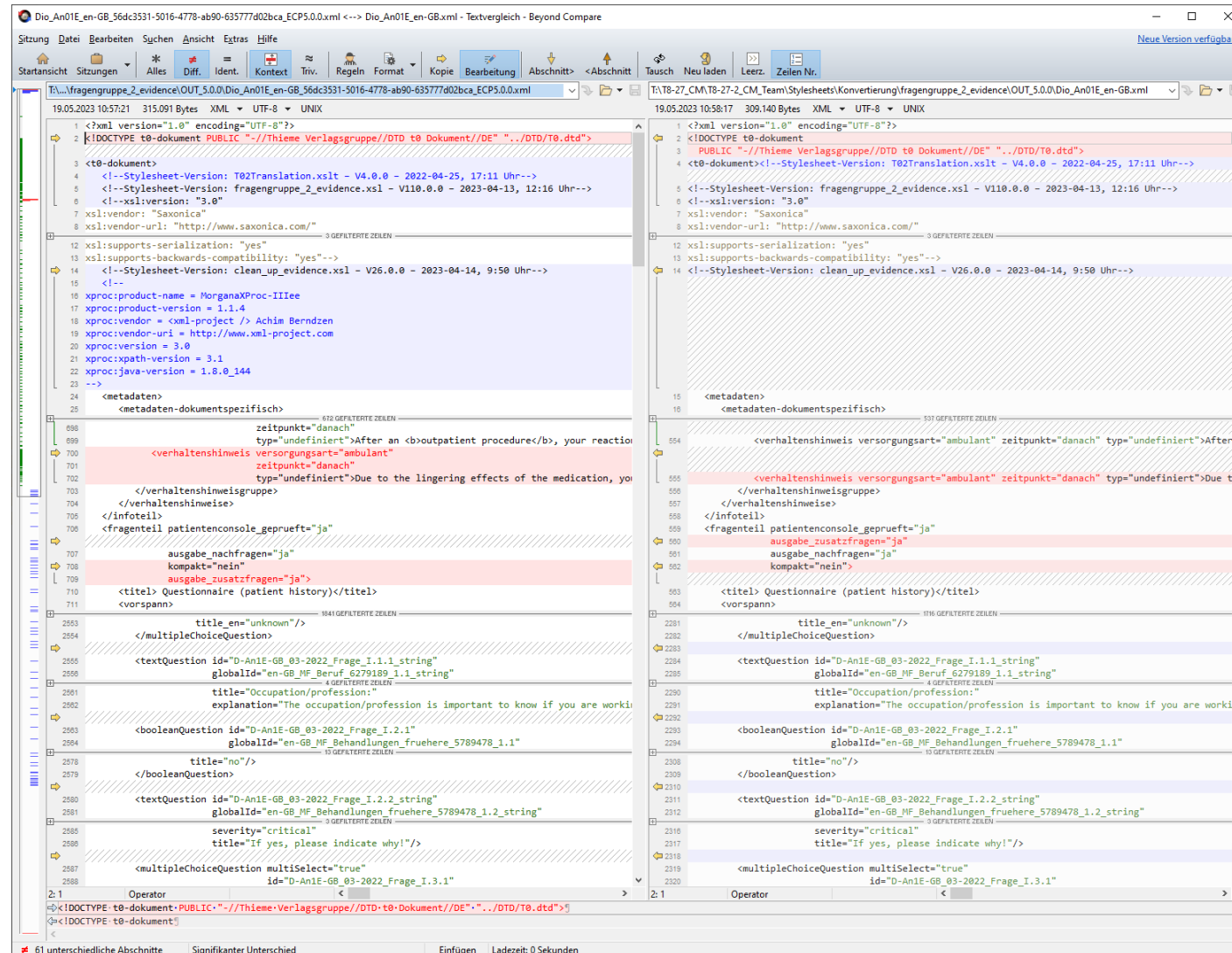
Takeaways

- The move from the batch scripts to XProc 3.0 was a smooth experience.
- XProc 3.0 proves to be a very good tool compared to XProc (1.0)
 - Syntactic sugar enables concise pipelines.
 - Cleaner step design prevents problems with document flow.
 - XPath 3.1 and XDM typing for options and variables are very helpful.
 - More powerful step libraries make processor specific steps unnecessary.
- MorganaXProc-IIIse turned out to be a reasonable tool:
 - The complete task could be fulfilled without any custom additions.
 - Some bugs had to be fixed to make the pipelines run.
 - Pain points for optimisation were identified and will be fixed in later releases.

Takeaways

- Comparing the results of the original batches with the pipeline results was difficult and took some time.

Takeaways: serialisation is now done by MorganaXProc and no longer by Saxon



... public ...

Takeaways

- Comparing the results of original batches with pipeline results was difficult and took some time.
- Low performance of the FHIR pipeline gave us some headaches.
 - Turned out that an alternative schema document can be used.
- XProc pipeline had to be optimised for acceptable performance, downgrading readability.

Takeaways

From

```
<p:for-each>  
  <p:xslt>  
    <p:with-input port="stylesheet" href="the-stylesheet.xml" />  
  </p:xslt>  
</p:for-each>
```

To

```
<p:load href="the-stylesheet.xml" name="stylesheet" />  
<!-- ... -->  
<p:for-each>  
  <p:xslt>  
    <p:with-input port="stylesheet" pipe="@stylesheet" />  
  </p:xslt>  
</p:for-each>
```


Takeaways

- Comparing the results of original batches with pipeline results was difficult and took some time.
- Low performance of the FHIR pipeline gave us some headaches.
 - Turned out that an alternative schema document can be used.
- XProc pipeline had to be optimised for acceptable performance, downgrading readability.
- Processor optimisation could be done, but it is not clear whether this conforms with the specs.
- Validating intermediate results with DTD is somewhat a pain because it is only possible on `<p:load>`.
- So please, XProc working group, support DTD validation natively and add `<p:validate-with-dtd>` and Bob's your uncle.

...:: public ::...



Thieme Compliance

Thank you for your attention!

Achim Berndzen, <xml-project />
achim.berndzen+xml-project.com



[@xml_project@mstdn.social](https://mstdn.social/@xml_project)

Thorsten Rohm, Thieme Compliance GmbH
thorsten.rohm@thieme-compliance.de



[linkedin.com/in/thorsten-rohm](https://www.linkedin.com/in/thorsten-rohm)



[@Markup_Fanatic@mastadon.cloud](https://mstdn.cloud/@Markup_Fanatic)



[@Markup_Fanatic](https://twitter.com/Markup_Fanatic)